

## 1. Úvod

JSTOR [1] (Journal Storage) je elektronický online archív, ktorý uchováva odborné časopisy, recenzie, zápisky z konferencií, monografie, a iného odborného obsahu, a zároveň poskytuje platformu pre bádanie najmä pre akademickú komunitu. Archív má bohatú podporu vydavateľov, knižníc aj inštitúcií kultúrneho dedičstva.

## 2. Vývoj archívu JSTOR

Vzhľadom na plniace sa zbierky univerzitných knižníc, vzrastajúce priestorové a finančné nároky na priestor a údržbu archívov, sa William G. Bowen, zakladateľ projektu JSTOR a bývalý prezident Univerzity Princeton rozhodol, že by bolo dobré na zníženie nákladov odborné časopisy zdigitalizovať a starosť o miesto a dostupnosť vďaka rozvíjajúcim sa informačným technológiám prenechať tretím stranám. [2]

JSTOR vzniká v roku 1995 nadáciou The Andrew W. Mellon Foundation a v prvý rok existencie bol dostupný pre 7 univerzít, rok neskôr sa jeho pole pôsobnosti rozšírilo o 10 ďalších. V súčasnej dobe je archív JSTOR dostupný pre vyše 6000 inštitúcií, z nich 800 tvoria stredné školy. V roku 1997 začínajú snahy sprístupniť archív online pomocou bežného webového browseru.

V juli roku 2008 nastáva zjednotenie JSTOR a ALUKA, digitálneho archívu zameraného na materiály pochádzajúce z Afriky alebo materiály o Afrike, ktorý je zastrešený organizáciou ITHAKA.

V januári roku 2009 dochádza k spojeniu JSTOR a organizácie ITHAKA. JSTOR sa stáva jednou zo služieb ITHAKA, čo je nezisková organizácia, ktorá pomáha akademickej komunite využívať digitálne technológie na zlepšenie bádania a učenia udržateľným smerom. Vznikla za podpory nadácií Mellon, Hewlett a Niarchos. [3]

V roku 2011 je plánované spustenie programu Current Scholarship Program.

## 3. Ciele projektu

V širšom zmysle cieľom projektu JSTOR je pomôcť akademickej komunite využiť pokrok v informačno – komunikačných technológiách. Aby tento cieľ dosiahol, JSTOR adoptoval systematicky širokú perspektívu, berúc do úvahy občas konfliktné potreby knižníc, vydavateľov a študentov – užívateľov.

JSTOR ciele zahŕňajú:

- vybudovať spoľahlivý a komplexný archív dôležitej akademickej časopiseckej literatúry,
- značne vylepšiť prístup k týmto časopisom,
- pomôcť zaplniť medzery v existujúcich knižničných kolekciami,
- venovať sa zachovaniu výtlačkov, ktoré sú buď zničené alebo opotrebované vekom,
- znížiť náklady knižníc spojené s priestorom a dlhodobým uchovávaním zbierok časopisov,
- pomôcť inštitúciám a vydavateľom prekenuť prechod k elektronickým spôsobom vydávania,

- študovať vplyv poskytovania elektronickej formy týchto odborných materiálov.

## 4. Digitalizácia

Štandardom pre digitalizáciu materiálu v archíve JSTOR zostáva viacstránkový TIFF vo verzii 6, pre obyčajný text sa používa bitonálny TIFF (tzn. bod má buď čiernu alebo bielu farbu) v rozlíšení 600 dpi. Kvôli obmedzeniu bitonálneho postupu sa pri nedostatočnej presnosti duplikácie digitálnej informácie použije buď čiernobiely alebo farebný TIFF v rozlíšení 300 dpi. Tieto postupy pri súčasnej úrovni techniky umožňujú tímu archívu JSTOR dosiahnuť priemernú presnosť pri OCR (*optical character recognition*) latinky, cyriliky aj gréckych písmen 99,5 %. [4]

## 5. Aktuálny stav archívu JSTOR

V súčasnej dobe sa počet časopisov uchovávaných archívom JSTOR zvýšil na 1151, čo zodpovedá 272678 zdigitalizovaným vydaniam a 2535833 kompletným článkom dostupným online. Do tohto čísla nie sú ešte zarátané úvodníky či recenzie kníh, ktoré zvyšujú počet dostupného materiálu na dvojnásobok. Medzi najstaršie materiály patria takmer 350 rokov staré výtlačky časopisu *The Philosophical Transactions of the Royal Society* z roku 1665, čo je najstarší tlačený časopis v anglicky hovoriacom svete. Jednotlivé články sú rozdelené v 51 rôznych disciplínach.[5]

Pre zníženie nákladov väčších odberateľov ako sú univerzity a iné vzdelávacie inštitúcie začal JSTOR poskytovať od roku 1997 časopisy rozdelené do 19 tematických kolekcíí. Tie sa ďalej delia na hlavné multidisciplinárne, vedľajšie oborovo špecifické kolekcie a kolekcie späté s bývalým archívom ALUKA. Multidisciplinárnych kolekcii je 10, a z toho ešte 3 čakajú na úplné zkompletovanie, ktoré má podľa plánu JSTOR prebehnúť ku koncu tohto a nasledujúcich 2 rokov. Napríklad Masarykova univerzita má predplatený jeden multidisciplinárny archív, a to Arts & Sciences I, a jeden oborovo špecifický, Business II.

JSTOR nie je databáza súčasných vydaní časopisov. Medzi najnovšie publikovanými vydaniami a vydaniami dostupnými v JSTOR je časový odstup – pohyblivá stena. Približne tri štvrtiny výtlačkov si udržuje pohyblivú stenu 3 alebo 5 rokov.

JSTOR pre podporu nezávislých výskumníkov umožňuje aj predaj článkov jednotlivcom prostredníctvom Publisher Sales Service. Takto avšak nie je možné zakúpiť každý, iba časopisy, u ktorých to povolil vydavateľ, ktorý zároveň získa z predaja vopred dohodnutý poplatok. V súčasnej dobe je pomocou služby Publisher Sales Service dostupných vyše 700 zbierok.

## 6. Práca s archívom JSTOR

V archíve sa pomocou metódy iteračného vyhľadávania hľadá ľahko, zjednodušenie prinášajú aj špeciálne vyhľadávacie znaky ako asterisk, wildcard, tilde, a i. (napr. ~ pre slová s podobným hláskovaním: vyhľadávacia otázka *dostoyevsky~* nájde výrazy *dostoyevsky*, *dostoevsky*, *dostoyevski*, *dostoevski*)

Vyhľadané články je však možné pre ďalšie použitie cez rozhranie archívu JSTOR uložiť iba vo formáte pdf, narozdiel od minulosti, keď bol materiál dostupný pre uloženie aj vo formáte TIFF, alebo je možné uložiť pomocou browseru obrázok textu, ktorý je vo formáte 16 bitového GIF, komprimovaný bezstratovou kompresiou LZW. Užívateľský komfort zlepšuje možnosť si po registrácii do systému zapamätať vyhľadané citácie, ktoré sa uchovávajú nehľadiac na to, cez akú inštitúciu je užívateľ práve na JSTOR pripojený.

JSTOR taktiež poskytuje jednoduché prepojenie so službou Google Scholar, kde možno vyhľadať príbuzné články k danému materiálu, články, ktoré citujú daný materiál, príp. ďalšie práce autora článku.

## 7. Current Scholarship Program

*Current Scholarship Program* (CSP) je nová snaha, započatá archívom JSTOR a vydavateľstvom Kalifornskej univerzity, vytvoriť súčasné i staršie publikácie dostupné na jednej, ucelenej platforme; poskytnúť jednotný bod pre knihovníkov aj koncových užívateľov po celom svete pre prístup k tomuto obsahu; a zaistiť dlhodobé uchovanie a dostupnosť daného obsahu.

Program začína rokom 2011. Vydavatelia, ktorí sa zapoja do CSP, budú mať oboje, súčasný i starší obsah z ich publikovaných časopisov dostupný na redizajnovanej JSTOR platforme. Cena súčasných vydaní bude nastavená daným účastniam sa vydavateľom a bude zahŕňať všetok obsah spreduhyblivej steny. Staršie vydania, ktoré nie sú v archíve JSTOR dostupné dnes, budú do archívu pridané.

Knižnice po celom svete si budú môcť licencovať súčasné vydania časopisov individuálne ako súčasť nového systému zbierok vydaní jednoducho pomocou označenia a pridania do už existujúcej licenčnej dohody. Tieto knižnice potom zaplatia archívu JSTOR jednotnú faktúru za všetok licencovaný obsah, vrátane už existujúcich starších zbierok za pohyblivou stenou.

Vydavatelia, ktorí sa zapoja do Programu, získajú pozornosť a moderné, cenovo prístupné prostriedky na spojenie sa s knižnicami a užívateľmi, pri rozumnej kontrole svojich cien, čo umožní lepšiu distribúciu pri spokojnosti oboch strán.

V októbri tohto roku sa do programu pridáva University of Illinois Press, ktorá publikuje niekoľko časopisov ako *American Journal of Psychology*, *American Music* či *Journal of Aesthetic Education*.

## 8. Data for Research

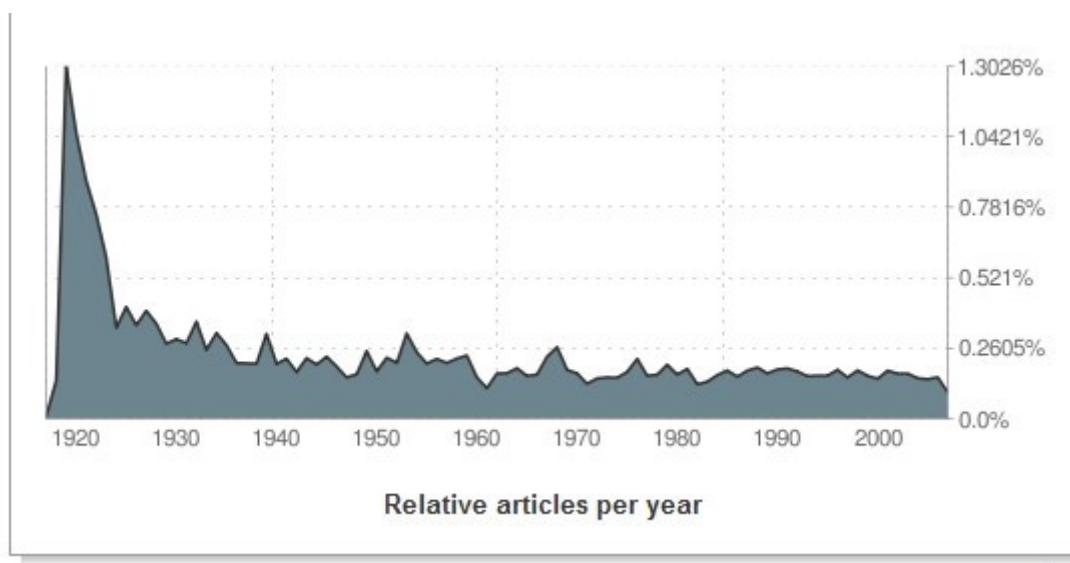
*Data for Research* (DfR) [6] je nová služba poskytovaná archívom. Ide skupinu webových aplikácií určených na interakciu s obsahom archívu JSTOR, ktorá má umožniť výskumníkovi iný spôsob vyhľadávania článkov a zároveň analyzovať získané dáta.

DfR umožňuje:

- Prehľadávať archív JSTOR pomocou pokročilého vyhľadávacieho rozhrania. Rýchlo pridať vyhľadávacie výrazy a filtrovať tak výsledky pre maximálnu presnosť.
- Pozrieť slovné informácie o vyhľadanom texte ako frekvencia výskytu slova, citácie, kľúčové pojmy či ngramy.
- Vyhľadať istú informáciu v danom obore, alebo priereze rôznych oborov. Nájsť najbežnejšie pojmy použité v konkrétnych oboroch.
- Sledovať frekvenciu použitia daných výrazov v rôznych časových obdobiach.
- Požiadat si o a stiahnuť dátové množiny archívu celého oboru pre vlastnú analýzu, dostupné však iba pre autorizované inštitúcie.

DfR realizuje prehľadávanie na viac než 5 miliónov článkov, čo zodpovedá 14 miliárd slov. Výsledky hľadania je možné filtrovať podľa času na časovej osi, ktorá zobrazuje graf relatívny počet článkov vydaných v danom roku, ktoré obsahujú hľadaný výraz. Množinu výsledkov je možné zužovať podľa ďalších kritérií, napr. vydavateľ, počet strán, použitie obrázkov, a i.

Sumár vyhľadávania je doplnený o jednoduchú grafickú reprezentáciu dát ako graf výskytu slova v článkoch rozdelený podľa oborov. Výsledky je možné stiahnuť vo formáte CVS.



Obrázok 1: (Reprezentácia článkov obsahujúcich slovo bolshevism. Je zreteľne vidieť, že daný výraz sa v odborných článkoch pred rokom 1917 prakticky nevyskytoval.)

DfR poskytuje aj API (*Applications Programming Interface*) pre programovateľné vyhľadávanie. Prehľadávanie sa dosahuje pomocou protokolu SRU (*Search and Retrieval*) cez URL a výsledok je vrátený vo formáte XML. Otázky môžu byť pokladané na 8 rôznych Dublin Core polí v metadátach (dc.title, dc.author, a i.) a 4 zvláštne jstor polia. (jstor.discipline, jstor.journaltitle) Dáta vrátené službou SRU obsahujú základné bibliografické údaje, vrátené ID článku je ďalej možno použiť na získanie počtu výskytu slova, ngramov, kľúčových pojmov alebo referencií.

Služba je momentálne v štádiu beta. Budúca funkcionálnosť má zahŕňať topic modeling a iné služby.

## 9. Vlastné zhodnotenie projektu

JSTOR si vytýčil cieľ uľahčiť inštitúciám prechod z klasickej na digitálnu formu archivácie diel. Vyše 6000 účastníacich sa inštitúcii napovedá tomu, že sa mu to darí. Pravdepodobne aj vďaka podpore poskytovanej tímom tvoriacim JSTOR, ktorý organizuje semináre, pracuje na nových službách, ale hlavne neustále doplňuje a rozširuje archív.

### Zoznam literatúry

- [1] JSTOR: Home. Dostupné z: <<http://www.jstor.org/>>.
- [2] JSTOR – Wikipedia, the free encyclopedia. Dostupné z: <<http://en.wikipedia.org/wiki/Jstor>>.
- [3] Announcement Archive. Dostupné z: <<http://www.jstor.org/page/info/about/news/announcements.jsp>>
- [4] Digitization Standards & Processes. Dostupné z: <<http://www.jstor.org/page/info/about/archives/process.jsp>>
- [5] JSTOR by the Numbers. Dostupné z <<http://www.jstor.org/page/info/about/archives/facts.jsp>>
- [6] Data for Research. Dostupné z: <<http://dfr.jstor.org/>>.

## Metadata v DC

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" lang="slo" content="JSTOR" />
<meta name="DC.Creator" content="Pavol Mihálik" />
<meta name="DC.Subject" content="JSTOR" />
<meta name="DC.Description" content="Zmapovanie doterajšieho vývoja, popísanie aktuálneho
stavu a nových aktivít rozvíjaných pod hlavičkou JSTOR – digitálneho archívu zameraného na
akademické časopisy." />
<meta name="DC.Date" content="2009-12-4" />
<meta name="DC.Type" content="esej" />
<meta name="DC.Format" content="application/pdf" />
<meta name="DC.Format" content="computerFile" />
<meta name="DC.Source" scheme="URL" content="http://www.jstor.org/" />
<meta name="DC.Language" content="sk" />
```